Can We Obtain Quantum Theory From Reasonable Axioms?

Lucien Hardy*
Centre for Quantum Computation,
The Clarendon Laboratory,
Parks road, Oxford OX1 3PU, UK

October 26, 2000

Abstract

The usual formulation of quantum theory is rather abstract (it employs complex Hilbert spaces, Hermitian operators, the trace rule, and unitary or superoperator evolution). It is natural to ask why the formalism is like this. In this paper we will show that, for finite dimensional systems, this rather abstract formalism follows from a set of reasonable axioms. These axioms might well have been postulated without reference to any experimental data. If one of the axioms is dropped then, rather than getting quantum theory, we get classical probability theory. In developing the axioms we obtain a representation of quantum theory entirely in terms of real numbers and, correspondingly, a generalization of the Bloch sphere to arbitrary dimension.

1 Introduction

Quantum theory, in its usual formalism, is very abstract. Pure states are represented by vectors in complex Hilbert spaces, or in the more general case of mixed states, by positive Hermitian operators. Measured quantities are represented by Hermitian operators and the probability of a particular outcome is given by the trace rule. The evolution of a state is given by a unitary operator or, more generally, by a superoperator (which may take a pure state to a mixed state). This formalism is the result of extensive interaction between experiment and theory. It is natural to ask why the formalism should be the way it is. There is, a tradition in physics of trying to obtain physical theories from pure thought alone without doing experiments. It is natural to wonder whether this is possible with quantum theory. Put another way, could a nineteenth century physicist have come up with quantum theory before the experimental data of the early twentieth century became available? In this paper we will show that the basic structure of quantum theory for finite dimension follows from a set of reasonable axioms that might have been posited by a nineteenth century physicist had he been troubled by one particular feature of classical probability theory. This feature is that, for a finite number of classical bits, it is not possible to have a continuous evolution from one pure state to another. In Axiom 5 below we assume that such continuous evolution is possible. When this axiom is dropped we get classical probability theory rather than quantum theory.

By the basic structure of quantum theory we mean that states and observables are represented by positive Hermitian operators, that measured probabilities are given by the trace rule, and that the evolution is as specified by quantum theory. We will not recover any particular form for the Hamiltonian since that belongs to particular applications of quantum theory (for example - a set of interacting spins, or the motion

*hardy@qubit.org

of a particle in one dimension).

In developing the axioms we will find a formulation of quantum theory entirely in terms of real vectors and real matrices. This gives us a generalization of the Bloch sphere to arbitrary dimension.

Various authors have set up axiomatic formulations of quantum theory [1, 2, 3, 4, 5, 6, 7, 8] (see also [9, 10]). The advantage of the present work is that there are a relatively small number of simple axioms, these axioms can easily be motivated without any particular appeal to experiment, and the mathematical methods required to obtain quantum theory from these axioms are very straightforward (essentially just linear algebra).

2 Sketch of paper

We will consider a situation in which a preparation device prepares a system. This system is allowed to pass through some device which may transform the state, and then it impinges on a measurement device. We assume that it takes K probability measurements (which we will call the fiducial measurements) to determine the state. The results of these probability measurements are arranged in a column vector $\mathbf{p}$ with kth entry equal to the kth measured probability. The nine axioms, which will be stated more precisely later, are

Axiom 0 Measured probabilities. *Measured probabilities are between 0 and 1.*

Axiom 1 Mixtures. *Probabilities for mixtures are given by the usual convex sum rule.*

Axiom 2 Basis. *There exists a set of N distinguishable states.*

Axiom 3 Exhaustiveness. *If a mathematical element is consistent with the other axioms then the corresponding physical element exists.*

Axiom 4 Mapping. *There exists a map from pure states to pure measurements which identify them (give probability equal to one) and this map takes straight lines to straight lines (so it conserves convex structure).*

Axiom 5 Continuity. *Pure states can be continuously and reversibly transformed into one another along a path through the pure states.*

Axiom 6 Simplicity. *There are no unnecessary degrees of freedom.*

Axiom 7 Uniformity. *The number of degrees of freedom associated with a subspace of a certain dimension is a function of the dimension.*

Axiom 8 Composite systems. *For a composite system consisting of subsystem A with (N_A, K_A) and subsystem B with (N_B, K_B), we have $N = N_A N_B$ and $K = K_A K_B$.*

Each of these axioms will be shown to be very well motivated without particular appeal to the specific detail of experimental data.

Quantum theory is reconstructed from these axioms in the following way. First we show that the state can be represented by a real vector $\mathbf{p}$. Then using Axiom 1 we show that a measurement of probability can be represented by the real vector $\mathbf{r}$, and the probability measured is given by

$$p_{\text{meas}} = \mathbf{r}.\mathbf{p} \tag{1}$$

We say that $\mathbf{p} \in S$ and $\mathbf{r} \in R$. We show that the sets S and R sets are convex. We define pure states and pure measurements to be the extremal normalized elements in these sets.

Using Axiom 4, we show that the state can be represented by a $\mathbf{r}$-type vector and that pure states satisfy an equation

$$\mathbf{r}^T D \mathbf{r} = 1 \tag{2}$$

where D is a $K \times K$ real matrix.

Next, Axioms 6, 7 and 8 are used to show that either $K = N$ or $K = N^2$. The $K = N$ case corresponds to the classical case. We use Axiom 5 to rule this out. This leaves $K = N^2$. First we consider the case $N = 2$ and $K = 4$. We show this case agrees with quantum theory for $N = 2$. In particular, we obtain the Bloch sphere for appropriate fiducial measurements. We then use this to obtain the form of states for general N.

Finally, we consider transformations. It follows from Axiom 1 that transformations are of the form

$$\mathbf{p} \rightarrow Z\mathbf{p} \tag{3}$$

where Z is a $K \times K$ real matrix. In quantum theory, the most general evolution is given by a superoperator \$

$$\rho \rightarrow \$(\rho) \tag{4}$$

We show that the constraints on Z that follow from the axioms correspond to the constraints on superoperators that follow from quantum theory.

3 Setting the scene

3.1 The setup

We will begin by describing the type of experimental situation we wish to consider (see Fig. 1). An experimentalist has three types of device. One is a preparation device. We can think of it as preparing physical systems in some state. It has on it a number of knobs which can be varied to change the state prepared. The system is then released by pressing a button. The system passes through the second device. This device can transform the state of the system. This device also has knobs on it which can be adjusted to effect different transformations (we might think of these as controlling fields which effect the system). We might allow the system to pass through a number of devices of this type. Unless otherwise stated, we will assume the transformation devices are set to allow the system through unchanged. Finally, we have a measurement apparatus. This also has knobs on it which can be adjusted to determine what measurement is being made. This device outputs a classical number. If the no system is incident on the device (i.e. because the button on the preparation device was not pressed) then it outputs a 0 (corresponding to a null outcome). If there is actually a physical system incident (i.e when the release button is pressed and the transforming device has not absorbed the system) then the device outputs a number l where $l = 1$ to L (we will call these non-null outcomes). The number of possible classical outputs, L, may depend on what is being measured (the settings of the knobs).

3.2 A measurement

We will consider only measurements of probability since all other measurements (such as expectation values) can be calculated from measurements of probability. When, in this paper, we refer to a *measurement* we mean, specifically, a measurement of the probability that the outcome belongs to some subset of the non-null outcomes with a given setting of the knob on the measurement apparatus. For example, we could measure the probability that the outcome is $l = 1$ or $l = 2$ with some given setting. To measure a probability we need a large number of identically prepared systems.

3.3 The state

The state of the system will be described by some yet to be specified mathematical object. To determine the values of the parameters in the state we need to make a number of different measurements on an ensemble of identically prepared systems. A certain number of appropriately chosen measurements will be both necessary and sufficient to determine the state. Let this number be K. Thus, for each setting, $k = 1$ to K, we will measure a probability p_k with an appropriate setting of the knob on the measurement apparatus. These probabilities can be represented by a column vector $\mathbf{p}$ where

$$\mathbf{p} = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ \vdots \\ p_K \end{pmatrix} \tag{5}$$

Now, this vector contains just sufficient information to determine the state and the state must contain just sufficient information to determine this vector (otherwise it could not be used to predict probabilities for measurements). In other words, the state and this vector are interchangeable and hence we can use $\mathbf{p}$ as a way of representing the state of the system. We

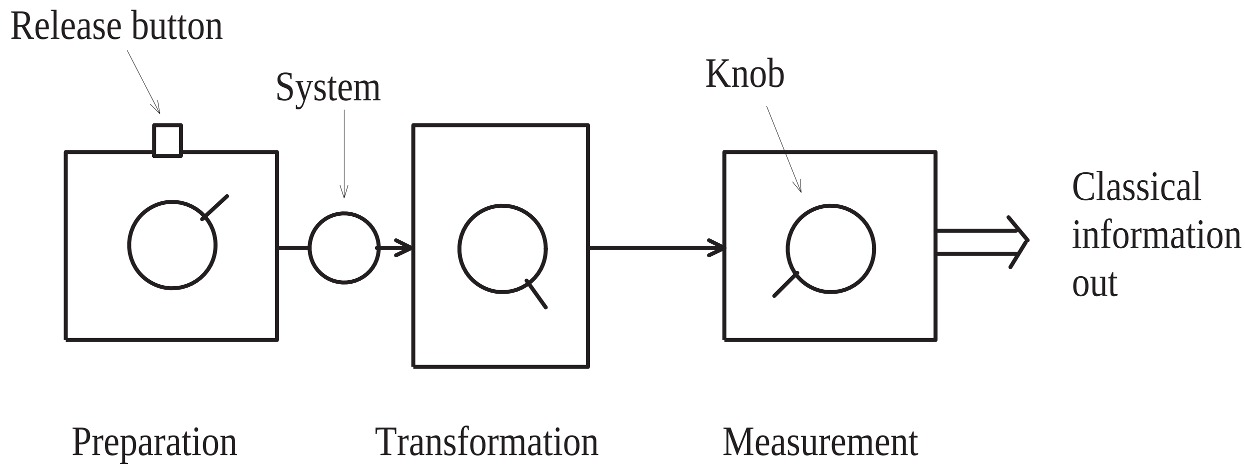

Figure 1: The situation considered consists of a preparation device with a knob for varying the state of the system produced and a release button for releasing the system, a transformation device for transforming the state (and a knob to vary this transformation), and a measuring apparatus for measuring the state (with a knob to vary what is measured) which outputs a classical number.

will call K the number of degrees of freedom associated with the physical system. We will not assume that the physical system is always present. Hence, one of the K degrees of freedom can be associated with normalization.

3.4 Fiducial measurements

We will call the probability measurements labeled by $k = 1$ to K used in determining the state the *fiducial* measurements. There is no reason to suppose that this set is unique. It is possible that some other fiducial set could also be used to determine the state.

4 Measured probabilities

Any probability that can be measured (not just the fiducial ones) will be determined by some function f of the state $\mathbf{p}$. Hence,

$$p_{\text{meas}} = f(\mathbf{p}) \tag{6}$$

For different measurements the function will, of course, be different. The first axiom is perhaps the most self-evident.

Axiom 0 *Measured probabilities are between 0 and 1.*

$$0 \leq p_{\text{meas}} \leq 1$$

This must be true if probabilities are measured by taking the proportion of cases in which a particular event happens in an ensemble.

4.1 Mixtures

The second axiom concerns mixtures of states. Assume that the preparation device is in the hands of Alice. She can decide randomly to prepare a state $\mathbf{p}_A$ with probability λ or a state $\mathbf{p}_B$ with probability $1 - \lambda$. Assume that she records this choice but does not tell the person, Bob say, performing the measurement. Let the state corresponding to this preparation be $\mathbf{p_C}$. Then, we expect that the probability Bob measures will be the the convex combination of the two cases, namely

$$f(\mathbf{p}_C) = \lambda f(\mathbf{p}_A) + (1 - \lambda) f(\mathbf{p}_B) \tag{7}$$

We will turn this into an axiom:

Axiom 1 Mixing states. *If a state C is prepared by preparing state A with probability λ and state B with probability $1-\lambda$ (and a record is kept but not revealed of which one has been prepared) then any measured probability on state C is given by λ times the probability had the state been A plus $1-\lambda$ times the probability had the state been B.*

This axiom is self-evident since Alice could subsequently reveal which state she had prepared for each event in the ensemble providing two sub-ensembles. Bob could then check his data was consistent for each subensemble. This could not be the case were axiom 1 not true.

4.2 Linearity

Axiom 1 can be applied to the fiducial measurements themselves. This gives

$$\mathbf{p}_C = \lambda \mathbf{p}_A + (1-\lambda)\mathbf{p}_B \tag{8}$$

This is clearly true since it is true by axiom 1 for each component.

Equations (7,8) give

$$f(\lambda \mathbf{p}_A + (1-\lambda)\mathbf{p}_B) = \lambda f(\mathbf{p}_A) + (1-\lambda) f(\mathbf{p}_B) \tag{9}$$

This strongly suggests that the function f is linear. This is indeed the case and a proof is given in Appendix 1. Hence, we can write

$$p_{\text{meas}} = \mathbf{r}\cdot\mathbf{p} \tag{10}$$

The vector $\mathbf{r}$ is associated with the measurement. The kth fiducial measurement is the measurement which picks out the kth component. Hence, the fiducial measurement vectors are

$$\mathbf{r}^1 = \begin{pmatrix}1\\0\\0\\\vdots\\0\end{pmatrix} \quad \mathbf{r}^2 = \begin{pmatrix}0\\1\\0\\\vdots\\0\end{pmatrix} \quad \mathbf{r}^3 = \begin{pmatrix}0\\0\\1\\\vdots\\0\end{pmatrix} \quad \text{etc.} \tag{11}$$

5 Transformations

We have discussed the role of the preparation device and the measurement apparatus. Now we will discuss the state transforming device (the middle box in Fig. 1). If some system with state $\mathbf{p}$ is incident on this device its state will be transformed to some new state $\mathbf{g}(\mathbf{p})$. It follows from the Axiom 1 that this transformation must be linear. This is clear since we can apply the proof in the Appendix 1 to each component of $\mathbf{g}$. Hence, we can write the effect of the transformation device as

$$\mathbf{p} \to Z\mathbf{p} \tag{12}$$

where Z is a $K \times K$ real matrix describing the effect of the transformation.

6 Exhaustiveness.

We now have three types of mathematical element representing states, measurements and transformations. These elements will belong to some set of physically allowed states, measurements and transformation. Let these sets of allowed elements be S, R and Γ. Thus,

$$\mathbf{p} \in S \tag{13}$$

$$\mathbf{r} \in R \tag{14}$$

$$Z \in \Gamma \tag{15}$$

We might find that there is a vector or matrix which does not belong to these sets but which is nevertheless consistent with all the other axioms. But in such a case there would be no reason for nature to exclude it. Therefore we will assume this does not happen.

Axiom 2 Exhaustiveness. *If a given vector (in the case of states and measurements) or a given matrix (in the case of transformations) is consistent with all the other axioms then there exists a corresponding physically realizable state, measurement, or transformation.*

This means that if there exists a vector $\mathbf{p} \notin S$, or a vector $\mathbf{r} \notin R$, or a matrix $Z \notin \Gamma$ but which is consistent with all the other axioms then the set S, or R or Γ must be expanded to include this as an element.

7 The nature of the set S of allowed states: Part A

7.1 S is convex

If $\mathbf{p}_A, \mathbf{p}_B \in S$ then it follow from (8) that the convex sum $\lambda \mathbf{p}_A + (1-\lambda)\mathbf{p}_B$ (for $0 \leq \lambda \leq 1$) is also in S. Hence, the set S is convex.

7.2 Normalized and unnormalized states

If the release button is never pressed then the fiducial measurements will all give zero probability. Hence, we can prepare the state represented by a vector $\mathbf{0}$ with zero for all its entries. Thus,

$$\mathbf{0} \in S \tag{16}$$

If the release button is always pressed (i.e for every event in the ensemble) then we will say $\mathbf{p} \in S_{\text{norm}}$ or, in words, that the state is normalized. Unnormalized states are of the form $\lambda \mathbf{p} + (1-\lambda)\mathbf{0}$ where $0 \leq \lambda < 1$.

7.3 Extremal states

S is a convex set. It is also bounded (by Axiom 0). Hence, it has an extremal set (namely the set of vectors in the set which cannot be written as a convex sum of pairs of other vectors in the set). Let S_{extremal} be the set of extremal states. The zero vector $\mathbf{0}$ is clearly a member of S_{extremal} since, by Axiom 0, it cannot be written as the convex sum of other vectors in S. Unnormalized states can be written as a convex sum of a normalized state and $\mathbf{0}$. Hence, they are not extremal states.

7.4 Pure states

We will define pure states as the set of extremal states except $\mathbf{0}$. All pure states are normalized (since all unnormalized states are not extremal). We will say that pure states are in S_{pure}. Pure states are clearly special in some way. They represent states which cannot be interpreted as a mixture. A driving intuition in formulating some of the remaining axioms is the idea that pure states represent definite (non-probabilistic) states of the system. We are familiar with such a concept in classical probability theory. For example, the 0 state of a classical bit is a definite state of a system.

8 The nature of the set R of allowed measurements: Part A

8.1 R is convex

If $\mathbf{r}_A, \mathbf{r}_B \in R$ then for $0 \leq \mu \leq 1$ we have

$$0 \leq (\mu \mathbf{r}_A + (1-\mu)\mathbf{r}_B) \cdot \mathbf{p} \leq 1 \tag{17}$$

Hence, we will have consistency with the axioms if the convex sum $\mu \mathbf{r}_A + (1-\mu)\mathbf{r}_B$ is in R and if therefore follows from Axiom 2 that it is. This means that the set R is convex. The physical interpretation of a convex sum of measurement vectors is that the actual measurement performed depends on the outcome of a coin toss.

8.2 The identity measurement

The probability of a non-null outcome is given by summing up all the non-null outcomes

$$p_{\text{non-null}} = \sum_{l=1}^{L} \mathbf{r}_l \cdot \mathbf{p} = \mathbf{r}^I \cdot \mathbf{p} \tag{18}$$

where $\mathbf{r}_l$ is the measurement vector corresponding to outcome l and

$$\mathbf{r}^I = \sum_{l=1}^{L} \mathbf{r}_l \tag{19}$$

is called the identity measurement.

8.3 Irreducible measurements

Consider a measurement vector $\mathbf{r} \in R$ which can be written as $\mathbf{r}_A + \mathbf{r}_B$ where the vectors $\mathbf{r}_A$ and $\mathbf{r}_B$ (not equal to $\mathbf{0}$ are in R and are non-parallel. We will say that such $\mathbf{r}$ are reducible. Let the set of measurement vectors which are not reducible be $R_{\text{irreducible}}$. Irreducible measurements are special because they cannot be regarded as composite measurements. Thus, recall that a measurement is when we measure the probability that the outcome belongs to some set of non-null outcomes. If there is more than one element in this set then the measurement is composite and reducible. On the other hand, if the measurement is irreducible then there is necessarily only one element in the set.

8.4 Normalization

If $\mathbf{r} \in R$ then it follows from Axiom 2 that $\mu\mathbf{r} \in R$ for $0 \leq \mu \leq 1$. It follows that a concept of normalization would be useful. We will define a way of normalizing measurement vectors after we have introduced the notion of a basis set.

9 Basis states

At the heart of classical probability theory is the idea that there exist a set of distinguishable states. For example, a classical bit has two such states, namely 0 and 1. We will introduce an axiom to maintain this idea. We require that there exist measurements which can be simultaneously performed (such as the $\mathbf{r}_l$ in equation (18). Further, we require that these measurements uniquely pick out the given state. Hence, we assume

Axiom 3 Basis set. *There exists a set of N states $\mathbf{p}_n$ and N measurement vectors $\mathbf{r}_n \in R_{\text{irreducible}}$ ($n = 1$ to N) which have the following properties (i) $\mathbf{p}_n$ is the only state for which $\mathbf{r}_n \cdot \mathbf{p} = 1$ (ii) $\mathbf{r}_m \cdot \mathbf{p}_n = \delta_{mn}$. (iii) $\sum_{m=1}^{N} \mathbf{r}_m = \mathbf{r}^I$.*

We impose that the measurement vectors be irreducible since we are not interested in composite measurements which have the basis property. Part (i) is necessary since we require that the measurement uniquely picks out the given state. Part (ii) is the usual basis property. The significance of part (iii) is that we require that all the measurements can be made at once (compare with equation (18)) so that the basis states can be distinguished for a single copy.

It follows from Axiom 3 that the basis states $\mathbf{p}_n$ are all pure. Assume the converse, that $\mathbf{p}_n = \lambda\mathbf{p}'_n + (1-\lambda)\mathbf{p}''_n$. It follows that $\mathbf{r}_n \cdot \mathbf{p}'_n = \mathbf{r}_n \cdot \mathbf{p}''_n = 1$. But, from (i) there can be only one vector with this property. Hence, $\mathbf{p}_n$ is pure.

9.1 The fiducial basis set

There may, in principle, exist different basis sets consistent with (i-iii) of Axiom 3. It is also possible that these basis sets have different N (though in fact this will turn out not to be the case). We will pick out a basis set having N equal to the smallest value it can take. We will call this basis set the fiducial basis set and we will call N the dimension.

10 The nature of R: Part B

10.1 Normalization

We will say that

$$\mathbf{r} \in R_{\text{norm}} \qquad \text{iff} \qquad \sum_{n=1}^{N} \mathbf{r} \cdot \mathbf{p}_n = 1 \tag{20}$$

where $\mathbf{p}_n$ are the states of the fiducial basis set. In other words, a measurement vector is normalized if we get a total of 1 when we add up the probabilities corresponding to that measurement for each state of the fiducial basis set.

We define a normalization coefficient for a measurement vector $\mathbf{r}_M$ to be

$$\nu = \mathbf{r} \cdot \mathbf{p}^I \qquad \text{where} \qquad \mathbf{p}^I = \sum_{n=1}^{N} \mathbf{p}_n \tag{21}$$

where $\mathbf{p}_n$ are the fiducial basis states. If $\nu = 1$ then $\mathbf{r} \in R_{\text{norm}}$. We can normalize a measurement vector by dividing it by its normalization coefficient.

10.2 Pure measurements

It is clear that R_{norm} is a convex set since (20) will be satisfied by any convex sum of any two states in R_{norm}. This convex set will have an extremal set. We will call this extremal set R_{pure}. Pure measurements are simply normalized irreducible measurements. Evidently the basis measurement vectors $\mathbf{r}_m$ introduced in Axiom 3 are pure (they are normalized since they satisfy (20) and they are irreducible by definition).

10.3 Generalized measurements

To help define generalized measurements we will first define a set R^+ which is bigger than R. We will say that $\mathbf{r} \in R^+$ iff there exists positive μ for which $\mu\mathbf{r} \in R$. Hence, $R \subset R^+$. It follows from the fact that a general vector in R will be a sum of irreducible vectors and that pure vectors are simply normalized irreducible vectors that a general vector in R^+ can be written as a sum of positive coefficients times pure measurement vectors. Hence, if we know R_{pure} then we can determine R^+. Once we know R^+ we only need some principle which restricts the length of the vectors to find R. Thus, consider consider an element of R^+. How do we know whether this is also an element of R? As long as it is in R^+ it cannot give rise to negative probabilities. Thus, we need to be sure that it will not give rise to probabilities greater than 1. It follows from equation (19) that there must exist another vector $\overline{\mathbf{r}} \in R$ such that $\mathbf{r} + \overline{\mathbf{r}} = \mathbf{r}^I$. Furthermore, if this is true then any measured probabilities must be less than or equal to 1. Hence,

$$\textbf{??}\text{Iff} \quad \mathbf{r}, \overline{\mathbf{r}} \in R^+ \quad \text{and} \quad \mathbf{r} + \overline{\mathbf{r}} = \mathbf{r}^I \quad \text{then} \quad \mathbf{r}, \overline{\mathbf{r}} \in R \tag{22}$$

since vectors satisfying this condition will not violate any of the axioms and therefore, by Axiom 2, must belong to R.

Note that it follows from (**??**) that $\mathbf{0} \in R$.

11 The nature of S: Part B

We define the normalization coefficient of a state $\mathbf{p}_S$ to be

$$\mu = \mathbf{r}^I \cdot \mathbf{p} \tag{23}$$

In the case where $\mathbf{p}_S \in S_{\text{norm}}$ we have $\mu = 1$.

If we choose the first N fiducial measurement vectors to be equal to the basis measurements, i.e.

$$\mathbf{r}^k = \mathbf{r}_k \quad \text{for} \quad k = 1 \quad \text{to} \quad N \tag{24}$$

then it follows from (11) and (19) that

$$\mathbf{r}^I = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{25}$$

where the first N entries are 1 and the remaining $K - N$ entries are 0. Hence, it follows from (23) that, for a normalized state, $\mathbf{p}$

$$\sum_{k=1}^{N} p_k = 1 \tag{26}$$

where it should be noted that the sum is only over the first N terms.

12 Alternative formulae for p_{meas}

We will label vectors pertaining to the system with subscript S and vectors pertaining to the measurement with subscript M. The state is represented by a vector $\mathbf{p}_S$ and a measurement is represented by a vector $\mathbf{r}_M$. We will now show that it is also possible to represent the state by a vector $\mathbf{r}_S$ and the measurement by a vector $\mathbf{p}_M$.

12.1 Fiducial states

We will choose K linearly independent states, $\mathbf{p}_S^k$ for $k = 1$ to K, and call them fiducial states (it must be possible to choose K linearly independent states since otherwise we would not need K fiducial measurements to determine the state). Consider a given measurement $\mathbf{r}_M$. We can write

$$p_M^k = \mathbf{r}_M \cdot \mathbf{p}_S^k \tag{27}$$

Now, we can take the number p_M^k to be the kth component of a vector. This vector, $\mathbf{p}_M$, is related to $\mathbf{r}_M$ by a linear transformation. Indeed, from the above equation we can write

$$\mathbf{p}_M = C\mathbf{r}_M \tag{28}$$

where C is a $K \times K$ matrix with l, k entry equal to the lth component of $\mathbf{p}^k$. Since the vectors $\mathbf{p}^l$ are linearly independent, the matrix C is invertible and so $\mathbf{r}_M$ can be determined from $\mathbf{p}_M$. This means that $\mathbf{p}_M$ is an alternative way of specifying the measurement. Since p_{meas} is linear in $\mathbf{r}_M$ which is linearly related to $\mathbf{p}_M$ it must also be linear in $\mathbf{p}_M$. Hence we can write

$$p_{meas} = \mathbf{p}_M \cdot \mathbf{r}_S \tag{29}$$

where the vector $\mathbf{r}_S$ is an alternative way of describing the state of the system. The kth fiducial state can be represented by an $\mathbf{r}$-type vector, $\mathbf{r}_S^k$, and is equal to that vector which picks out the kth component of $\mathbf{p}_M$. Hence, the fiducial states are

$$\mathbf{r}_S^1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \qquad \mathbf{r}_S^2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \qquad \mathbf{r}_S^3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} \qquad \text{etc.} \tag{30}$$

12.2 More linear forms for p_{meas}

The expression for p_{meas} is linear in both $\mathbf{r}_M$ and $\mathbf{r}_S$. In other words, it is a bilinear form and can be written

$$p_{\text{meas}} = \mathbf{r}_M^T D \mathbf{r}_S \tag{31}$$

where superscript T denotes transpose, and D is a $K \times K$ real matrix (equal, in fact, to C^T). The k, l element of D is equal to the probability measured when the kth fiducial measurement is performed with the lth fiducial state (since, in the fiducial cases, the $\mathbf{r}$ vectors have one 1 and otherwise 0's as components). Hence,

$$D_{lk} = (\mathbf{r}_M^l)^T D \mathbf{r}_S^k \tag{32}$$

The expression for p_{meas} must also be a bilinear form in $\mathbf{p}_M$ and $\mathbf{p}_S$. Hence, we can also write

$$p_{\text{meas}} = \mathbf{p}_M^T G \mathbf{p}_S \tag{33}$$

where G is another $K \times K$ dimensional matrix. In fact, by comparing (10,29,31,33), we can see that $G = D^{-1}$ (D is invertible since the fiducial set of states are linearly independent).

12.3 Vectors associated with states and measurements

There are two ways of describing the state: Either with a $\mathbf{p}$-type vector or with an $\mathbf{r}$-type vector. From (10, 31) we see that the relation between these two types of description is given by

$$\mathbf{p}_S = D\mathbf{r}_S \tag{34}$$

Similarly, there are two ways of describing the measurement: Either with an $\mathbf{r}$-type vector or with a $\mathbf{p}$-type vector. From (29,31) we see that the relation between the two ways of describing a measurement is

$$\mathbf{p}_M = D^T \mathbf{r}_M \tag{35}$$

(Hence, C in equation (28) is equal to D^T.)

13 Measurements which identify states

We will say that the measurement $\mathbf{r}_M$ *identifies* the state $\mathbf{p}_S$ iff $\mathbf{r}_M \cdot \mathbf{p}_S = 1$, i.e. if we get a measured probability equal to 1 when that measurement is performed on that state. The identity measurement identifies all normalized states.

The basis measurement $\mathbf{r}_n$ identifies the basis state $\mathbf{p}_n$. In this case both the measurement and the state are pure. More generally, we expect there to be a map which takes pure states to those pure measurements which identify them. This maps between S_{pure} and R_{pure}. But these sets consist only of extremal points. Hence, we want to be sure that the map has the property that it never maps three points in S_{pure} which are, by definition, not on a straight line, to three points on a straight line since these points could not then be in R_{pure}. The only map which will do this irrespective of the set S_{pure} is of the form $\mathbf{r} = M\mathbf{p}+\mathbf{a}$ where M is a $K \times K$ real invertible matrix and $\mathbf{a}$ is a constant vector. We would also expect the map to take $\mathbf{0} \in S$ to $\mathbf{0} \in R$ since these are extremal elements, which implies that $\mathbf{a} = \mathbf{0}$. Hence, we will assume

Axiom 4 State to measurement map. *There exists a map of the form*

$$\mathbf{r} = M\mathbf{p},$$

where M is a $K \times K$ real matrix, which maps any pure state, $\mathbf{p}$, to a pure measurement, $\mathbf{r}$, which identifies that pure state.

It will follow below that the matrix M is invertible so we need not assume this. It is not logically necessary that this axiom be satisfied. Thus, if the map was nonlinear, then there may exist an extremal set which is mapped to another extremal set. However, this would require a very particular choice of extremal set. The map here has the added advantages that it preserves the convex structure inside the sets S and R and that it puts these sets on a symmetric footing.

13.1 Choosing fiducial states and fiducial measurements

A convex structure embedded in a K-dimensional space must have at least $K + 1$ extremal points (for example, a triangle has three extremal points, a tetrahedron has four, etc.). In the case of the set S, one of these extremal points will be $\mathbf{0}$ leaving at least K remaining extremal points which will correspond to pure states (recall that pure states are extremal states other than $\mathbf{0}$). Furthermore, it must be possible to choose a set of K of these pure states which correspond to linearly independent vectors (if this were not possible then the convex hull would be embedded in a lower than K dimensional space). Hence, we can choose all our fiducial states to be pure. Let these fiducial states be $\mathbf{p}_S^k$. We will choose the kth fiducial measurement $\mathbf{r}_M^k$ to be that pure measurement which uniquely identifies the kth fiducial state (this measurement exists by Axiom 4). Hence,

$$(\mathbf{r}_M^k) \cdot \mathbf{p}_S^k = 1 \tag{36}$$

for the fiducial vectors. The state of the system can be represented by $\mathbf{r}_S$ rather than $\mathbf{p}_S$, these vectors being related by $\mathbf{p}_S = D\mathbf{r}_S$. Corresponding to the set S (of $\mathbf{p}_S$) will be a set Q (containing $\mathbf{r}_S$) given by acting on each element of S with D^{-1}. The sets Q_{extremal}, Q_{norm} and Q_{pure} can be similarly defined. From Axiom 4 there must exist a map between R and Q (since Q is linearly related to S) of the form

$$\mathbf{r}_S = H\mathbf{r}_M \tag{37}$$

where H is a $K \times K$ matrix which maps between pure states and the pure measurements which identify them. In particular this is true for the fiducial states and measurements:

$$\mathbf{r}_S^k = H\mathbf{r}_M^k \tag{38}$$

However, the fiducial vectors have the special form given in (11,30), namely zeros everywhere except for the kth entry. Hence, the map H is equal to the identity. This is true because we have chosen the fiducial measurements to be those which identify the fiducial states. Since Q and R are related by the identity map we will drop the M and S subscripts in what follows, it being understood that the leftmost vector corresponds to the measurement apparatus and the rightmost vector corresponds to the state. Thus we now have that $\mathbf{r}$ identifies $\mathbf{r}$ if $\mathbf{r}$ is pure. Hence,

$$\mathbf{r}^T D\mathbf{r} = 1 \tag{39}$$

for pure states. This equation is very useful since will enable us to find the pure states.

14 The continuity axiom

Now we introduce the axiom which will give us quantum theory rather than classical probability theory. Consider a pure state. These are the states that have a definite property - namely that there exists a pure measurement with definite outcome. We expect to be able to apply reversible transformations on a system which will evolve these states into one another (by reversible we mean that the effect of the transforming device (the middle box in Fig. 1.) can be reversed irrespective of the input state and hence that Z^{-1} exists). Thus, consider a transformation between two pure states:

$$\text{pure state A} \leftrightarrow \text{pure state B}$$

We expect that any such transformation be continuous since there are generally no discontinuities in physics. Hence, unless these two pure states are represented by infinitesimally close vectors, the state must pass through some intermediate state.

$$\text{pure state A} \leftrightarrow \text{intermediate state} \leftrightarrow \text{pure state B}$$

We expect this intermediate state to be pure. Were it not pure then we would be going from a state with definite properties, throwing away some information to obtain the impure intermediate state and then regaining information to obtain a pure state again. Then, when we reverse the evolution we would obtain the original pure state again even though we apparently threw information away. These considerations motivate the next axiom.

Axiom 5 Continuity. *There exists a continuous and reversible transformation between any two pure states along a continuous trajectory through the pure states.*

We can see immediately that classical systems of finite dimension N will run into problems with the continuity part of this axiom. Consider, for example, transforming a classical bit from the state 0 to the state 1. Any continuous transformation would have to go through an infinite number of other pure states (not part of our finite system). Indeed, this is clear given any physical implementation of a classical bit. For example, a ball in one of two boxes must move along a continuous path from one box (representing a 0) to the other box (representing a 1). Deutsch has pointed out that for this reason, the classical description is necessarily approximate in such situations whereas the quantum description in the analogous situation is not approximate [11]. We will use this axiom to rule out various theories which do not correspond to quantum theory (including classical probability theory).

This axiom can be further motivated by thinking about computers. A classical computer will only employ a finite number distinguishable states (usually referred to as the memory of the computer - for example 10Gbytes). For this reason it is normally said that the computer operates with finite resources. However, if we demand that these bits are described classically and that transformations are continuous then we have to invoke the existence of an infinite number of distinguishable states for the reasons stated above. Hence, the resources used by a classically described computer performing a finite calculation must be infinite. It would seem extravagant of nature to employ infinite resources in performing a finite calculation.

15 Subspaces

Consider a system of dimension N. From Axiom 3 this has a basis set consisting of N vectors, $\mathbf{r}_n$, satisfying

$$\mathbf{r}^I = \sum_{n=1}^{N} \mathbf{r}_n \tag{40}$$

We could consider a subset of $N' < N$ of these basis vectors. We expect that there will be some sense in which degrees of freedom can be associated with this subset forming a subspace and that the degrees of freedom associated with this subspace will have the same properties as those of a system simply having dimension N'. Let us formalize these notions. Let W be a set with N' members consisting of distinct integers between 1 and N. Define

$$\mathbf{r}^I_W = \sum_{n\in W} \mathbf{r}_n \tag{41}$$

This vector plays the same role as $\mathbf{r}^I$ but in the subspace associated with this subset of basis vectors. Now consider those state vectors $\mathbf{p}$ which have support only with respect to this subset of basis vectors. We define the set S_W by

$$\mathbf{p} \in S_W \quad \text{iff} \quad \mathbf{r}_W^I \cdot \mathbf{p} = \mathbf{r}^I \cdot \mathbf{p} \tag{42}$$

State vectors in S_W will be said to belong to the subspace S_W defined with respect to the basis vectors $\mathbf{r}_n$ with $n \in W$. Note that it follows from this definition that the set S_W is convex. We will say that this subspace is of dimension N' (though this is may be a little misleading since there may be more than N' degrees of freedom associated with the subspace).

15.1 Fiducial subspaces

There may exist many different basis sets. We have selected one - the fiducial basis set. We will call subspaces defined with respect to the fiducial basis set *fiducial subspaces*.

16 The fiducial states and subspaces

We can chose the first N of the K fiducial states to be the fiducial basis states

$$\mathbf{p}^k = \mathbf{p}_k \qquad \text{for} \quad k = 1 \quad \text{to} \quad N \tag{43}$$

Then

$$\mathbf{p}^k \in S_{W=\{k\}} \qquad \text{for} \quad k = 1 \quad \text{to} \quad N \tag{44}$$

In this case the fiducial subspace $S_{W=\{k\}}$ is of dimension 1 since W has one member. What about the remaining $K - N$ fiducial states. They cannot be in any of the sets $S_{W=\{k\}}$ since these the fiducial states can be taken to be normalized and since, by part (i) of Axiom 3, $\mathbf{p}_k$ is the only normalized state in $S_{W=\{k\}}$. Consider $\mathbf{p}^{N+1}$. Maybe

$$\mathbf{p}^{N+1} \in S_{W=\{1,2,3\}} \tag{45}$$

In this case the fiducial subspace S_W is of dimension 3. Or maybe

$$\mathbf{p}^{N+1} \in S_{W=\{1,2\}} \tag{46}$$

In this case fiducial subspace S_W is of dimension 2. In general, all the fiducial vectors will belong to some fiducial subspace. Now there will be many ways of choosing the fiducial states. Assume we adopt the following method. First, we choose as many of the fiducial states as possible to belong to 1 dimensional fiducial subspaces (so W has only one element - these fiducial states will simply be the N fiducial basis states). Then, after this is done, we choose as many of the remaining fiducial states as possible to belong to 2 dimensional fiducial subspaces. If there are still fiducial states left over after this then we continue choosing as many of these remaining fiducial states as possible to belong to 3 dimensional fiducial subspaces, and so on. Let us say that, when proceeding in this way, the last fiducial state goes into a fiducial subspace of dimension $s_{\max}$. We will assume that $s_{\max}$ takes the maximum value consistent with the other axioms.

Axiom 6 No unnecessary degrees of freedom. *If the fiducial states are chosen so that as many as possible belong to one dimensional fiducial subspaces, then, of the remainder, as many as possible belong to two dimensional fiducial subspaces, and so on, up to $s_{\max}$ dimensional fiducial subspaces, then we assume that $s_{\max}$ takes the minimum value consistent with the other axioms.*

The motivation behind this axiom is as follows. In principle there is no limit to the number of degrees of freedom K for a given dimension N. Therefore, it seems reasonable that nature will seek to be efficient and have as small a number of degrees of freedom as possible. We can imagine starting with small dimensional systems $N = 1, 2, \ldots$ and building up. We would like to introduce as few new degrees of freedom as possible as we build up to higher dimensions. Axiom 6 captures exactly this idea.

A subspace of dimension N_W can be used to simulate a system actually having this dimension. We expect a certain uniformity on the part of nature. Thus, if a system, or subsystem has dimension 5 say then we expect that it will have some given number of degrees of freedom, e.g. 25. We expect that the number of degrees of freedom will be independent of whether we have a subspace, a full space, and independent of the way in which the system was prepared. Hence, we assume,

Axiom 7 Uniformity. *The number of degrees of freedom, K_W, associated with a N_W dimensional subspace is given by a fixed function of N_W, namely*

$$K_W = K(N_W)$$

In particular, applied to the whole system, we have $K = K(N)$. We would like to find this functional form. First we will introduce the final axiom.

17 Composite systems

It often happens that a preparation device ejects its system in such a way that it can be regarded as being made up of two subsystems. For example, it may emit one system to the left and one to the right (see Fig. 2). We will label these subsystems A and B. We assume

Axiom 8 Composite systems. *A system consisting of two composite subsystems A and B having dimension N_A and N_B respectively, and number of degrees of freedom K_A and K_B respectively, has dimension $N = N_A N_B$ and number of degrees of freedom $K = K_A K_B$.*

We expect this to be true for the following reasons. If the subsystems A and B have N_A and N_B distinguishable states, then we would expect there to be $N_A N_B$ distinguishable states for the whole system. To see why the relationship $K = K_A K_B$ is reasonable consider first two uncorrelated subsystems. A likely possible choice of fiducial measurements for this system is where we perform the ith fiducial measurement on system A and the j fiducial measurement on system B and measure the joint probability p_{ij} that both measurements have a positive outcome. The state $\mathbf{p}$ would then consist of all these numbers arranged in a column vector. For uncorrelated subsystems these joint probabilities are equal to $p_{iA}p_{jB}$. Thus, although there are $K_A K_B$ such joint probabilities, they are specified by only $K_A + K_B$ real numbers for uncorrelated subsystems. However, we must also be able to consider convex sums of such preparations. In this case, the joint probabilities will not factorize. Instead they will have the form

$$p_{ij} = \int h(\lambda) p_{iA}(\lambda) p_{jB}(\lambda) d\lambda \tag{47}$$

where h is a probability density satisfying

$$\int h(\lambda) = 1 \tag{48}$$

and $h \geq 0$. Now we have $K_A K_B$ numbers p_{ij}. These numbers are independent. The reason for this is that although the numbers are constrained to be of the form (47), these constraints correspond to inequalities (essentially Bell type inequalities). Hence, they do not eliminate degrees of freedom but rather define a volume of dimension $K_A K_B$ in the space spanned by these probabilities [12]. Hence we will need $K_A K_B$ fiducial measurements to determine the position of the state in this space. It is possible that, when two systems are combined, extra degrees of freedom come into existence. Axiom 8 says that this does not happen.

It should be emphasized that it is not required by the axioms that the vector $\mathbf{p}$ have components satisfying (47). Indeed, it is the fact that there can exist vectors not of this form that leads to quantum entanglement.

18 The relationship between N and K

We will now apply Axioms 6, 7 and 8 to find the relationship between N and K. Axiom 6 says that $s_{\max}$ is minimum. Thus, first we try $s_{\max} = 1$. Consider the 1 dimensional fiducial subspace with $W = \{n\}$. There are N such subspaces. Associated with each degree of freedom is a fiducial state. We can choose all our fiducial states to be normalized. It follows from part (i) of Axiom 3 that $\mathbf{p}_n$ is the only normalized vector in this subspace. There are N such 1 dimensional subspaces and each must have one fiducial state which gives us N fiducial states. Since we are assuming $s_{\max} = 1$ we have counted all the fiducial states so we have

$$K = N \quad \text{for} \quad s_{\max} = 1 \tag{49}$$

We will show below that the $K = N$ case (which corresponds to classical probability) violates the continuity axiom and hence we must consider $s_{\max} = 2$. Again, we have N fiducial states contributed from the 1 dimensional fiducial subspaces. Now consider the 2 dimensional fiducial subspaces. By Axiom 7 each of these must have some constant number of degrees of freedom $K_2 = K(2)$. Of these K_2 degrees of freedom two have already been counted since they belong to one dimensional subspaces. This leaves $K_2 - 2$ degrees of freedom. There are a total of $N(N-1)/2$ two dimensional fiducial subspaces (corresponding to the number of ways two basis vectors can be selected from N). It follows from part (i) of Axiom 3 that a pair of distinct two dimensional fiducial subspaces can have, at most, only one vector in common, this being a shared fiducial basis state (if these subspaces do happen to have one fiducial basis vector in common). Hence, it follows that these $K_2 - 2$ additional states are independent for each of the $N(N-1)/2$ possible two dimensional fiducial subspaces. Hence the total number of degrees of freedom is

$$K = N + 12N(N-1)(K_2 - 2) \quad \text{for} \quad s_{\max} = 2 \tag{50}$$

Axioms 7 and 8 imply

$$K(N_A N_B) = K(N_A)K(N_B). \tag{51}$$

We can solve equations (50) and (51) for K_2. We obtain

$$K_2 = 2, 4 \tag{52}$$

Substituting these back into (50) we find that

$$K = N \qquad \text{or} \qquad K = N^2 \tag{53}$$

We know already that $K = N$ corresponds to $s_{\max} = 1$ which will be shown to be inconsistent with the continuity axiom. It will turn out that the $K = N^2$ case is consistent with all the other axioms. Hence, by axiom 6, we need go no further than $s_{\max} = 2$. Furthermore, we will show that the $K = N^2$ case corresponds to the quantum theory.

We could consider higher values of $s_{\max}$. For example, when $s_{\max} = 3$ then we obtain a solution in which there is one degree of freedom state associated with each 1 dimensional fiducial subspace, 6 degrees of freedom in addition to the 2 already counted associated with each 2 dimensional fiducial subspace, and 6 degrees of freedom in addition to those already counted associated with each 3 dimensional subspace. It is not clear, however, that solutions for higher values of $s_{\max}$ can be made consistent with the other axioms (in particular that they can correspond to a convex set).

19 $K = N$: Classical probability theory

Consider first the $K = N$ case. There will be $K = N$ fiducial vectors which we can choose to be equal to the basis vectors. From equation (32) we know that the lk element of D is equal to the measured probability with the kth fiducial state and the lth fiducial measurement. Since the fiducial vectors correspond to basis vectors this implies that D is equal to the identity. The pure vectors $\mathbf{r} \in Q_{\text{pure}}, R_{\text{pure}}$ must satisfy

$$\mathbf{r}^T D \mathbf{r} = 1 \tag{54}$$

We also have $\mathbf{p} = D\mathbf{r}$ (equation (34)). Given that D is equal to the identity in this case we obtain

$$\sum_{k=1}^{N} (p^k)^2 = 1 \tag{55}$$

where p^k is the kth component of $\mathbf{p}$. However, we also know that

$$0 \leq p^k \leq 1 \tag{56}$$

We must also impose normalization using equation (26). The solutions of (26), (55) and (56) are when one p_k is equal to 1 and all the others are equal to 0. In other words, the only pure vectors are the basis vectors themselves. This forms a discrete set of vectors and so it is impossible for Axiom 5 (the continuity axiom) to be satisfied. Hence, we rule out such theories. However, we note that this is exactly classical probability theory. The pure vectors plus the zero vector represent definite states. A general state is a convex sum of these states and has a unique decomposition in terms of these definite states. The coefficient p^k can be regarded as the probability that the system is in the kth basis state. If the state is normalized (so there are no null events) then we have $\sum_{k=1}^{N} p^k = 1$ as we expect in classical probability theory. The convex set will be a polytope. For example, if $N = 2$ then it will be a rightangled triangle with the vertices represent the null event, $(0, 0)$, and the two definite outcomes, $(1, 0)$ and $(0, 1)$.

20 $K = N^2$: Quantum theory

20.1 $N = 2$ and $K = 4$

Having ruled out the $K = N$ case we will now consider the $K = N^2$ case. We will start by considering the simplest non-trivial example, namely that with $N = 2$ and $K = 4$. We can take the first two fiducial vectors to be basis vectors. The remaining two fiducial vectors will remain unspecified, though, as described in section 13.1, we will choose them to be pure and such that the kth fiducial measurement identifies the kth fiducial state (which implies that the diagonal elements of D will be equal to 1). Hence, the D matrix has the form

$$D = \begin{pmatrix} 1 & 0 & 1-a & 1-b \\ 0 & 1 & a & b \\ 1-a' & a' & 1 & c \\ 1-b' & b' & c' & 1 \end{pmatrix} \tag{57}$$

The reason for the 0's is that the first two fiducial vectors are basis vectors. The a and $1-a$ pair (and the other similar pairs) follow from normalization. Consider two vectors $\mathbf{r}$ and $\mathbf{r}'$. Let these vectors be normalized to μ and ν respectively. We can use these normalization coefficients to eliminate one of the degrees of freedom in the vectors. This is done in Appendix 2. It is shown that

$$p_{\rm meas} = \mathbf{r}^T D \mathbf{r}' = \vec{v}^T A \vec{v}\,' + \mu\nu/2 \tag{58}$$

where

$$\vec{v} = \begin{pmatrix} r_2 - r_1 \\ r_3 \\ r_4 \end{pmatrix} \tag{59}$$

and

$$A = \begin{pmatrix} 12 & a-12 & b-12 \\ a'-12 & 12 & c-12 \\ b'-12 & c'-12 & 12 \end{pmatrix} \tag{60}$$

In Appendix 3 it is shown, using Axiom 5, that the matrix A is symmetric. This implies immediately that D is symmetric for $N = 2$ and, as we will see below, that D is symmetric in general for all N. Thus, we must put $a' = a$, $b' = b$, and $c' = c$. It is significant that we have been able to derive this symmetry property. The corresponding symmetry property is assumed by some authors setting up axiomatic schemes for quantum theory [5, 6, 7].

All the pure states will be normalized. Furthermore, they will satisfy $\mathbf{r}^T D \mathbf{r} = 1$ or

$$\vec{v}^T A \vec{v} = 12 \tag{61}$$

This equation defines a two dimensional surface T embedded in three dimensions. For example, if $a = b = c = 12$ then we have a sphere of radius 1 (this is, in fact, the Bloch sphere). If A has three positive eigenvalues then T will be an ellipsoid. If A has one or two negative eigenvalue then T will be a hyperboloid. A hyperboloid is not a convex surface and so cannot represent the points in the set of pure vectors (it is not necessary that all the points on the hyperboloid correspond to physical states so a more careful statement is that we cannot choose a continuous subset of points on the hyperboloid which form the extremal surface of a convex set and which include all the fiducial states). Thus we require that T have three positive eigenvalues (if A has three negative eigenvalues then there cannot be any real solutions for $\vec{v}$). A necessary condition for A to have all positive eigenvalues is that $\det(A) > 0$. We have three variables a, b and c. The condition $\det(A) = 0$ is satisfied when

$$c = c_\pm \equiv 1 - a - b + 2ab \pm 2\sqrt{ab(1-a)(1-b)} \tag{62}$$

Note, we get the same conditions on c if we solve $\det D = 0$. We know the case with $a = b = c = 1/2$ corresponds to a sphere. This falls between the two roots in equation (62). The sign of the eigenvalues cannot change unless the determinant passes through a root as the parameters are varied. Hence, all values of a, b, c satisfying

$$c_- < c < c_+ \tag{63}$$

must correspond to three positive eigenvalues and hence to an ellipsoid. Values outside this range correspond to some negative eigenvalues (this can be checked by trying a few values). Hence, (63) must be satisfied. We will see below that this is exactly the condition that must be satisfied by quantum theory.

If we put $a = b = c = 12$ then

$$D = \begin{pmatrix} 1 & 0 & 1/2 & 1/2 \\ 0 & 1 & 1/2 & 1/2 \\ 1/2 & 1/2 & 1 & 1/2 \\ 1/2 & 1/2 & 1/2 & 1 \end{pmatrix} \tag{64}$$

$$A = \begin{pmatrix} 1/2 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/2 \end{pmatrix} \tag{65}$$

and the ellipsoid T takes the simplest form. It is, in fact, then a sphere of radius 1 satisfying

$$\vec{v} \cdot \vec{v} = 1 \tag{66}$$

We will work with this simple form below. The other forms correspond to choosing different fiducial vectors. They are, however, physically equivalent. This will be clear when we show that all ellipsoid forms of W consistent with (63) correspond to a choice of quantum fiducial vectors. They are then, all physically equivalent and hence there is no loss of generality in working with the simple spherical case.

20.2 The Bloch sphere normalized to μ

If the vector $\mathbf{r}$ is normalized to μ then $\mathbf{r}^T D\mathbf{r} \leq \mu^2$, or, from equations (58) and (65)

$$\vec{v} \cdot \vec{v} \leq \mu^2 \tag{67}$$

Hence, all vectors normalized to μ lie on or inside a sphere of radius μ. Those vectors lying on the surface of the sphere are proportional to a pure vector. To specify the position of a vector on the surface we need only specify the spherical coordinates (θ, ϕ). Thus, for a vector on the surface we have

$$r_2 - r_1 = \mu \cos\theta \tag{68}$$

Using $\mathbf{p} = D\mathbf{r}$ and (64) we obtain

$$p_2 - p_1 = r_2 - r_1 \tag{69}$$

Since $\mu = p_1 + p_2$ we obtain finally

$$p_1 - p_2 p_1 + p_2 = \cos\theta \tag{70}$$

Hence, $\cos\theta$ fixes the ratio p_2/p_1. If θ and the normalization coefficient μ is known then this fixes p_1 and p_2. The phase angle ϕ then fixes the values of p_3 and p_4.

Opposite points on a Bloch-sphere correspond to a basis for the 2 dimensional subspace. The fiducial basis is is given by $\vec{v_1} = (-1, 0, 0)$ and $\vec{v_2} = (1, 0, 0)$.

It is interesting to note that the Bloch sphere normalized to one maps to a sphere of radius 12 nested just inside the unit cube with one vertex at the origin in the first octant of the the space spanned by p_2, p_3, and p_4.

20.3 The general N case

For general N and K the state will be represented by a vector with K entries. Pure states will be solutions to the equation $\mathbf{r}^T D\mathbf{r} = 1$. We could attempt a procedure similar to that adopted in the $N = 2$ case above. However, there is a simpler method which also provides greater insight into quantum theory.

Consider the vectors $\mathbf{p}$ and $\mathbf{r}$. These have K entries, the kth of which is associated with the kth fiducial vector. The first N entries are associated with the basis states. Hence, $\sum_{k=1}^{N} p_k = \mu$ where μ is the normalization coefficient. The remaining entries are associated with 2 dimensional fiducial subspaces. Consider the two dimensional fiducial subspace $S_{W=\{i,j\}}$. Associated with this will be the ith and jth entries of the vectors $\mathbf{p}$ and $\mathbf{r}$ plus two other entries. We can take these four entries, cross out the remaining entries, and form the *truncated* vectors $\underline{\mathbf{p}}$ and $\underline{\mathbf{r}}$ corresponding to this fiducial subspace. The sum of the first two elements of the truncated vector $\underline{\mathbf{p}}$ is equal to the normalization coefficient μ_{ij} for this subspace (since they correspond to basis measurements). We can also truncate the $K \times K$ matrix D to give the 4×4 matrix $\underline{D}$ corresponding to this subspace. This matrix will be of the form (64) above. The same reasoning follows for this subspace as followed in the $N = 2$ case. Hence, we can say that any vector $\underline{\mathbf{r}}$ normalized to μ_{ij} must satisfy a Bloch sphere equation $\vec{v} \cdot \vec{v} \leq \mu_{ij}^2$. Put simply, there is a μ_{ij}-normalized Bloch sphere associated with each two dimensional fiducial subspce.

An extremal vector must be extremal up to normalization in each subspace. Hence, it follows that if the state is pure then the Bloch vector $\vec{v}$ is at the surface of the μ_{ij}-normalized Bloch sphere for each of the $N(N-1)/2$ two dimensional fiducial subspaces. Thus, the position of the pure vector will be specified by specifying all the pairs (θ_{ij}, ϕ_{ij}) for each 2 dimensional fiducial subspace. However, as we will now see, this over specifies a pure state. Instead, it is sufficient to specify only the $2N - 2$ real parameters (p_n, ϕ_{1n}) for $n = 2$ to N. To prove this we need only show that from these $2N - 2$ real parameters we can calculate all the pairs (θ_{ij}, ϕ_{ij}).

First, note that since the state is pure it must be normalized. Hence, $\sum_{n=1}^{N} p_n = 1$. If we know p_n for $n = 2$ to N then we can calculate p_1. Then we can calculate $\cos\theta_{ij}$ (and hence θ_{ij} since it is in the range 0 to π) for each 2 dimensional fiducial subspace using (70).

Now it remains to calculate the ϕ_{ij}. Consider basis measurements $\mathbf{r}_1$, $\mathbf{r}_i$ and $\mathbf{r}_j$ for $i \neq j$. Associated with these are the probabilities p_1, p_i and p_j. Put $p_1 + p_i + p_j = \mu_{1ij}$. Note that if we know μ_{1ij} and if we know θ_{1i} and θ_{1j} then we can use equation (70) to determine p_1, p_i and p_j. We can then use (70) to calculate θ_{ij}. This is not very useful since we already know θ_{ij}. But we can use this trick to determine ϕ_{ij} by applying it in rotated subspaces. Thus, consider a rotation in the $S_{\{1,i\}}$ subspace to new basis vectors $\mathbf{r}'_1$, $\mathbf{r}'_i$. Correspondingly, we can rotate in the $S_{\{1,j\}}$ subspace so that the new basis vectors are $\mathbf{r}'_1$, $\mathbf{r}'_j$ ($\mathbf{r}'_1$ is necessarily in common with these two subspaces). In doing this we have also effected a rotation in the $S_{\{i,j\}}$ subspace to the new basis vectors $\mathbf{r}'_i$ and $\mathbf{r}'_j$. Corresponding to the three new primed basis states will be the primed probabilities p'_1, p'_i and p'_j which will still add to μ_{1ij} (since these rotations do not take us out of this 3 dimensional subspace) and the primed angles θ'_{1i} and θ'_{1j}. We know μ_{1ij} since we know all the p_n. We can calculate θ'_{1i} and θ'_{1j} since we know (θ_{1i}, ϕ_{1i}) and (θ_{1j}, ϕ_{1j}) and we know the parameters of the rotations. From θ'_{1i} and θ'_{1j} and μ_{1ij} we can calculate the primed probabilities and from these probabilities we can calculate θ'_{ij}. There will be some equation relating primed and unprimed spherical coordinates. Hence, if we know $\theta' ij$ for sufficiently many different rotations then we can calculate ϕ_{ij}. This completes the proof.

From (p_n, ϕ_{1n}) for $n = 2$ to N we can calculate the Bloch vector in the two dimensional fiducial subspaces, and hence we can calculate $\mathbf{r}$ for the whole space. In this way we can determine the pure states and pure measurements as parameterized by the $2N - 2$ real numbers. Once we have the pure states we know all the states since they are the convex sum of pure states and the zero vector. Similarly, once we have the pure measurements we can determine all the measurements applying (Rplus).

The convex shape we obtain here is embedded in an N^2 dimensional space. If we assume the states are normalized then it is embedded in a $N^2 - 1$ dimensional space. The surface of this shape will have $N^2 - 2$ dimensions. However, the pure states live in only a $2N - 2$ dimensional space. In the particular case $N = 2$ we have a 3 dimensional convex shape for normalized states with a 2 dimensional surface and a 2 dimensional extremal set. In this case the pure states coincide with the the surface. It is this case we are most familiar with since this is the Bloch sphere. However, this is a very special case. In general, the pure states will live in a space of much lower dimension than the surface of the normalized convex set of states. For $N = 3$ we have 8 dimensional convex shape for normalized states. The surface of this is 7 dimensional. However, the pure states live in a 4 dimensional space.

There is an alternative approach to finding the pure vectors for the general $K = N^2$ case. We can use the $N = 2$ case to construct D in the general case. This is easy. The $K \times K$ matrix D will have mostly zeros. The only non-zero entries in D are when the i and j position correspond to fiducial vectors in the same 2 dimensional fiducial subspace. In these cases D_{ij} can be read off from (64). Once we have D we can solve for pure vectors using $\mathbf{r}^T D\mathbf{r} = 1$ and normalization. The corresponding equation is the natural generalization of the Bloch sphere to higher dimensions. This will be seen to correspond to quantum theory since we will get the same D matrix in that case.

21 This is quantum theory

21.1 States

In quantum theory we have

$$p_{\text{meas}} = \text{tr}(\widehat{A}\widehat{\rho}) \tag{71}$$

where $\widehat{A}$ is a positive operator associated with the measurement and $\widehat{\rho}$ is the density operator describing the state. We can put this in terms of real numbers. First, choose a linearly independent set of K projection operators $\widehat{P}^k$ for $k = 1$ to K. We will call these the fiducial projection operators. Since $\widehat{A}$ and $\widehat{\rho}$ are Hermitian we can write

$$\widehat{A} = \sum_{k=1}^{K} r_M^k \widehat{P}^k \equiv \mathbf{r}_M \cdot \widehat{\mathbf{P}} \tag{72}$$

and

$$\widehat{\rho} = \sum_{k=1}^{K} r_S^k \widehat{P}^k \equiv \widehat{\mathbf{P}} \cdot \mathbf{r}_S \tag{73}$$

where r_M^k and r_S^k are real, and $\widehat{\mathbf{P}}$ is the column vector with kth entry equal to $\widehat{P}^k$. These two equations can be inserted into (71) giving

$$p_{\text{meas}} = \sum_{ij} r_M^i \text{tr}(\widehat{P}^i \widehat{P}^j) r_S^j \tag{74}$$

which we can write as

$$p_{\text{meas}} = \mathbf{r}_M^T D \mathbf{r}_S \tag{75}$$

where the matrix D has entries

$$D_{ij} = \text{tr}(\widehat{P}^i \widehat{P}^j) \tag{76}$$

We can use $\mathbf{p} = D\mathbf{r}$ to obtain

$$\widehat{A} = \mathbf{p}_M^T D \widehat{\mathbf{P}} \tag{77}$$

and

$$\widehat{\rho} = \mathbf{P}^T D \mathbf{p}_S \tag{78}$$

Hence, we can interconvert between the standard quantum description of the state and the measurement operators and the description used in this paper in terms of real vectors.

Consider a density matrix $\widehat{\rho}$ corresponding to a N dimensional system. It requires N^2 real numbers to specify this Hermitian matrix (there are N real numbers along the diagonal, and $N(N-1)/2$ complex numbers above the diagonal). To determine these N^2 real parameters will require N^2 probability measurements. Hence, $K = N^2$ as required. Now consider the $N = 2$, $K = 4$ case. Let the 4 fiducial projection operators be

$$\widehat{P}^1 = |0\rangle\langle 0| \tag{79}$$

$$\widehat{P}^2 = |1\rangle\langle 1| \tag{80}$$

$$\widehat{P}^3 = (\alpha|0\rangle + \beta|1\rangle)(\alpha^*\langle 0| + \beta\langle 1|) \tag{81}$$

$$\widehat{P}^3 = (\gamma|0\rangle + \delta^*|1\rangle)(\gamma^*\langle 0| + \delta^*\langle 1|) \tag{82}$$

where $|\alpha|^2 + |\beta|^2 = 1$ and $|\gamma|^2 + |\delta|^2 = 1$. We can calculate D using (76)

$$D = \begin{pmatrix} 1 & 0 & 1-|\beta|^2 & 1-|\delta|^2 \\ 0 & 1 & |\beta|^2 & |\delta|^2 \\ 1-|\beta|^2 & |\beta|^2 & 1 & |\alpha\gamma^* + \beta\delta^*|^2 \\ 1-|\delta|^2 & |\delta|^2 & |\alpha\gamma^* + \beta\delta^*|^2 & 1 \end{pmatrix} \tag{83}$$

This has the same form as (57) above if we put $a = a' = |\beta|^2$, $b = b' = |\delta|^2$ and $c = c' = |\alpha\gamma^* + \beta\delta^*|^2$. By varying the complex phase associated with α, β, γ

and δ we find that $c_- < c < c_+$ where $c_\pm$ are defined in equation (62). This constraint is exactly the same constraint on the matrix D as the constraint derived from the axioms. Hence, we see that for the $N = 2$ case the axioms give quantum states.

For higher dimension, we can construct the D matrix from (76). This will give us the same D matrix we would construct from the method from the axioms as described above. Hence we obtain the same pure states (the solutions to $\mathbf{r}^T D\mathbf{r} = 1$). Indeed, we know that pure states in quantum theory can be written as

$$|\psi\rangle = \sqrt{p_1}|1\rangle + \sqrt{p_2}e^{i\phi_{12}}|2\rangle + \sqrt{p_3}e^{i\phi_{13}}|3\rangle + \dots \tag{84}$$

This state is specified by the same $2N - 2$ real parameters we found earlier from the axioms.

21.2 measurements

The positive operator $\widehat{A}$ is part of a positive operator valued measure (POVM). In a generalized quantum measurement each outcome $l = 1$ to L will be associated with a positive operator $\widehat{A}_l$ such that the measured probability of the lth outcome is given by $\mathrm{tr}(\widehat{A}_l\widehat{\rho})$. These operators need only satisfy the constraints that they are positive (and therefore Hermitian) and that

$$\sum_{l=1}^{L} \widehat{A}_l = \widehat{I} \tag{85}$$

where $\widehat{I}$ is the identity operator. We can see immediately that (19) gives (85). Since $\widehat{A}_l = \widehat{\mathbf{P}} \cdot \mathbf{r}_l$ we have

$$\sum_{l=1}^{L} \widehat{A}_l = \sum_{l=1}^{L} \widehat{\mathbf{P}} \cdot \mathbf{r}_l = \widehat{\mathbf{P}} \cdot \mathbf{r}^I \tag{86}$$

Now, if we choose the first N fiducial projection operators to form a basis then $\sum_{k=1}^{N} \widehat{P}^k = I$. Correspondingly, we can choose the first N fiducial measurements to be fiducial basis measurements. Hence, $\mathbf{r}^I$ has the form given in (25). Thus, the right hand side of (86) is equal to $\widehat{I}$ as required. The other constraint on $\widehat{A}$ is that it is positive. An operator will be positive iff it can be written as the sum of positive coefficients times projection operators. This is equivalent to demanding that the measurement vector $\mathbf{r}$ belong to R^+ (see section 10.3) which is a necessary condition on the measurement vectors that follows from the axioms.

We have shown that the theory of states and measurements we derived from the axioms is the same as quantum theory. All that remains to be considered are the transformations.

22 Conditions on transformations

In this section we will see that the conditions on transformations that can be derived from the axioms are the same as the conditions on transformations that derive from quantum theory.

22.1 Transformations for composite systems

To do this we need to develop a little more carefully what the axioms say for composite systems. Fig. 2. shows a preparation apparatus producing a system made up of subsystems A and B such that A goes to the left and B goes to the right. These subsystems then impinge on measurement apparatuses after passing through transformations devices which perform transformations Z_A and Z_B. This set up can be understood to be a special case of the more generic setup shown in Fig. 1. (there is no stipulation in the case of Fig. 1. that the measurement apparatus or any of the other apparatuses be located only in one place). Assume the transformation devices are initially set to leave the subsystems unchanged. From Axiom 8 we know that there are $K_A K_B$ fiducial measurements. One choice of fiducial measurements is where we simply perform the ith fiducial measurement on A and the jth fiducial measurement on B and measure the joint probability p_{ij}. The probabilities p_{ij} can be put in the form of a column vector $\mathbf{p}_{AB}$. However, for discussing transformations, it is more convenient to put them in the form of a

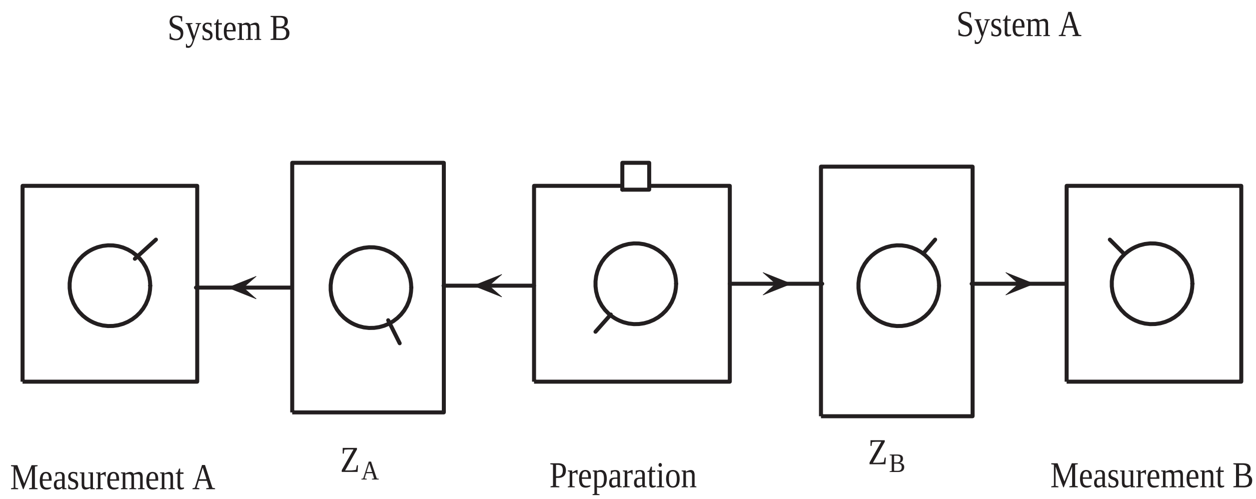

Figure 2: The preparation device here prepares a system in the form of two subsystems which go to the left and the right.

$K_A \times K_B$ matrix, $\widetilde{p}_{AB}$, having ij entry p_{ij}. It is easy to convert between these two ways of describing the state. We could regard both the preparation apparatus and measurement apparatus B as a preparation apparatus preparing states of subsystem A. If we perform the jth fiducial measurement on system B and take only those cases where we obtain a positive result for this measurement then the resulting state of system A will be given by a vector proportional to the jth column of $\widetilde{p}_{AB}$ (since these probabilities are proportional to the probabilities that would be obtained for the fiducial measurements on A with this preparation). Hence, the columns of $\widetilde{p}_{AB}$ must transform under Z_A. Similarly, the rows of $\widetilde{p}_{AB}$ must transform under Z_B. Hence, when the transformation devices in Fig. 2. are active, we have

$$\widetilde{p}_{AB} \rightarrow Z_A \widetilde{p}_{AB} Z_B^T \tag{87}$$

We will say that Z_A is completely positive iff

$$\widetilde{p}_{AB} \rightarrow Z_A \widetilde{p}_{AB} \tag{88}$$

maps all allowed $\mathbf{p}_{AB}$ to allowed $\mathbf{p}_{AB}$ for any dimension K_B.

22.2 Conditions on transformations

The only constraint on transformation matrices Z is that they transform states in S to states in S. This means that probabilities must remain bounded by 0 and 1. Hence,

1. Z must not increase the normalization coefficient of states.

2. Z must be completely positive.

Condition 2 is necessary since any system could always be a subsystem of some larger system.

In quantum theory, the most general evolution of $\widehat{\rho}$ is under a superoperator \$

$$\widehat{\rho} \rightarrow \$(\widehat{\rho}) \tag{89}$$

It has been shown by Kraus [13] that the following constraints are sufficient to restrict \$ to the cases allowed by quantum theory.

A \$ is linear

B \$ preserves hermiticity

C \$ does not increase trace

D \$ is completely positive

Complete positivity of $\$_A$ means that $\$_A \otimes I_B$ is positive (maps positive operators to positive operators) for any dimension of system B.

We will now see that each of these conditions is satisfied by the transformations Z. We have

$$\widehat{\rho} = \widehat{\mathbf{P}} \cdot \mathbf{r} \rightarrow \$(\mathbf{P}) \cdot \mathbf{r} \tag{90}$$

If \$ is linear then

$$\$(\widehat{P}^k) = \sum_j Y_{jk} \widehat{P}^j \tag{91}$$

Hence,

$$\mathbf{r} \rightarrow Y\mathbf{r} \tag{92}$$

where Y is the matrix with entries Y_{jk}. Hence, a linear transformation on $\widehat{\rho}$ corresponds to a linear transformation on $\mathbf{r}$ (and therefore on $\mathbf{p}$). Since Z effects a linear transformation of $\mathbf{p}$ it satisfies condition A. Condition B is satisfied by Z since it is real and therefore $\widehat{\rho}$ remains Hermitian. Condition C is the same as condition 1 above and condition D is the same as condition 2. Hence, the most general transformations allowed by the axioms correspond to the most general transformations allowed by quantum theory.

23 Conclusions

We have shown how quantum theory, for finite dimensional systems, can be obtained from a set of reasonable axioms. If Axiom 5 is dropped then we get classical probability theory instead. A by-product of this work is a formulation of quantum theory entirely in terms of real numbers and, correspondingly, a generalization of the Bloch sphere approach to arbitrary dimensions.

If a physicist in the 19th century had been dissatisfied with continuity problems in classical probability theory, he might have tried to develop a more satisfactory theory. It is conceivable that the alternative theory he developed would have been quantum theory. All the axioms used here can be motivated without appealing to any particular experimental data.

Though these results hold for finite dimension, a countable infinity of dimensions can be understood as a limiting case. However, for continuous infinite dimensional spaces it is not so clear what would happen. We may expect to be able to apply the axioms to any finite dimensional subspace and hence recover quantum theory that way. However, continuous spaces are notoriously difficult to deal with and hence a careful analysis is required.

From this work it is clear that quantum theory is, in some respects, better than classical probability theory. It can describe continuous evolution in a system with only a finite number (or countable infinity) of distinguishable states. If we have reason to believe that the universe has only a countable number of distinguishable states then we could not use classical probability theory. One possible reason is that we expect the continuous description of space to break down at the Plank scale.

It is curious that complex numbers drop out of this analysis. We might also expect to get a solution where pure states are represented by a vector in a real Hilbert space. We can use quantum theory to simulate such a space (for example, linearly polarised photons). We can even derive Bell's theorem using only such vectors. However, the axioms here insist on giving us the full richness of a complex Hilbert space.

There are many reasons to look for better axiomatic formulations of quantum theory.

- Pure aesthetics. We can are more likely to believe a theory that follows from reasonable axioms.

- By having a set of reasonable axioms we can look for ways in which we violate these axioms in a still reasonable way in the hope of going beyond quantum theory (for example, to develop quantum gravity).

- This approach puts a different slant on the interpretation of quantum theory (see discussion below).

- Since the formulation of quantum theory here is closer to classical probability theory than the standard formulation, this may motivate new applications and new treatments of the theory of quantum information.

There are various ways in which this work has a bearing on interpretational matters. First, if we really believe these axioms to be reasonable then they would apply to any hidden variables. If these hidden variables can only take a finite number distinguishable states then it would follow that any hidden variable substructure must look like quantum theory. Thus, we could not use hidden variables to solve the measurement problem (since this relies on being able to give the hidden variables a classical probability interpretation). The alternative is that hidden variables are allowed to take a continuous infinity of distinguishable values. But then we would be using infinite computing resources for finite dimensional quantum systems. Second, we see here how successful a purely instrumentalist approach is in obtaining the structure of quantum theory. This need not contradict beliefs held by the realist since he would anyway expect quantum theory to be consistent with instrumentalist argumentation. And, third, we obtain that the most general evolution is that of a superoperator. This is capable of taking pure states to mixed states. Hence, collapse interpretations of quantum theory could be incorporated into this structure.

Acknowledgements

I am very grateful to Chris Fuchs for discussions that motivated this work and to Jeremy Butterfield, Philip Pearle and Jos Uffink for comments. This work is funded by a Royal Society University Research Fellowship.

References

[1] G. Birkhoff and J. von Neumann, Ann. Math. **37**, 743 (1936).

[2] G. W. Mackey, *The mathematicalfoundations of quantum mechanics* (W. A. Benjamin Inc, New York, 1963).

[3] J. M. Jaunch and C. Piron, Helv. Phys. Acta **36**, 837 (1963); C. Piron, Helv. Phys. Acta **37**, 439 (1964).

[4] G. Ludwig, Commun. Math. Phys. **9**, 1 (1968), G. Ludwig, *Foundations of quantum mechanics* volumes I and II (Springer-Verlag, New York, 1983 and 1985).

[5] B. Mielnik, Commun. Math. Phys. **9**, 55 (1968).

[6] A. Lande, Am. J. Phys. **42**, 459 (1974).

[7] D. I. Fivel, Phys. Rev. A **50** 2108 (1994).

[8] B. Coecke, D. Moore, A. Wilce, *Current research in operational quantum logic: algebras, tategories, languages* (Fundamental theories of physics series, Kluwer Academic Publishers, 2000), also available on quant-ph/0008019.

[9] A. M. Gleason, Annals of Math **6**, 885 (1957).

[10] S. Kochen and E.P. Specher, J. Math and Mech. **17**, 59 (1967).

[11] D. Deutsch, private communication.

[12] I. Pitowsky, Lecture notes in physics **321** (Springer-Verlag, Berlin-Heildelburg 1989).

[13] K. Kraus, *States, effects, and operations: Fundamental notions of quantum theory* (Springer-Verlag, Berlin, 1983); B. Schumacher, quant-ph/9604023 (appendix A); J. Preskill *Lecture notes for physics 229: quantum information and computation*, available at http://ww.theory.ca.tech.edu/ preskill/ph229 (see chapter 3).

Appendix 1

We wish to prove that the convexity axiom implies

$$f(\alpha \mathbf{p}_A + \beta \mathbf{p}_B) = \alpha f(\mathbf{p}_A) + \beta f(\mathbf{p}_B) \qquad (93)$$

if

$$\mathbf{p}_A, \mathbf{p}_B, \alpha \mathbf{p}_A + \beta \mathbf{p}_B \in S$$

for all α, β where S is the set of allowed $\mathbf{p}$. First note that putting $\mathbf{p}_A = \mathbf{0}$ (it is shown in Section 7.2 that $\mathbf{0} \in S$) in equation (9) gives

$$f(\lambda \mathbf{p}) = \lambda f(\mathbf{p}) \tag{94}$$

for $0 \leq \lambda \leq 1$. We can write $\gamma = 1/\lambda$ and $\mathbf{p}'' = \mathbf{p}/\lambda$. Then we obtain

$$f(\gamma \mathbf{p}'') = \gamma f(\mathbf{p}'') \tag{95}$$

where $1 \leq \gamma$. Hence,

$$f(\mu \mathbf{p}) = \mu f(\mathbf{p}) \tag{96}$$

if $\mu \geq 0$. This only follows from convexity if $\mathbf{p}, \mu\mathbf{p} \in S$. However, if this is not the case, then the equation does not correspond to any physical situation. Hence, we are free to impose that (96) is true for all $\mathbf{p}$. In those cases where $\mathbf{p}, \mu\mathbf{p} \in S$ is not satisfied the equation has no physical significance anyway.

It is possible that one of the fiducial measurements is that measurement which simply checks that the system is present (i.e. that a null event is not seen). Let this measurement correspond to the first component of $\mathbf{p}$. If the first component of $\mathbf{p}_{A,B}$ is a, b then the corresponding normalized states are $\tilde{\mathbf{p}}_{A,B}$ are defined through $a\tilde{\mathbf{p}}_A = \mathbf{p}_A$ and $b\tilde{\mathbf{p}}_B = \mathbf{p}_B$. The first component of $\alpha\mathbf{p}_A + \beta\mathbf{p}_B$ is $\alpha a + \beta b$ and we require that this is between 0 and 1 (since this is equal to the normalization factor).

Now consider the case where α and β are both positive. Then we can write

$$\alpha\mathbf{p}_A + \beta\mathbf{p}_B = \alpha a\tilde{\mathbf{p}}_A + \beta b\tilde{\mathbf{p}}_B + (1 - \alpha a + \beta b)\mathbf{0}$$

The coefficients in this expression are all positive and add to 1. Hence, we can apply convexity. Since $f(\mathbf{0}) = 0$ we obtain (93).

Now consider the case where α is positive but β is negative. Put

$$\gamma\tilde{\mathbf{p}}_C = \alpha\mathbf{p}_A + \beta\mathbf{p}_B$$

where γ is chosen such that $\tilde{\mathbf{p}}_C$ is normalized and hence $\gamma = \alpha a - |\beta| b$. Since $\alpha\mathbf{p}_A + \beta\mathbf{p}_B$ must be in S we require that $\alpha a - |\beta| b$ is between 0 and 1. We can rearrange the above vector equation to obtain

$$\alpha a - |\beta| b\alpha a\tilde{\mathbf{p}}_C + |\beta| b\alpha a\tilde{\mathbf{p}}_B = \tilde{\mathbf{p}}_A \tag{97}$$

The coefficients on the LHS are both positive and add to 1. Hence convexity gives

$$\alpha a - |\beta| b\alpha a f(\tilde{\mathbf{p}}_C) + |\beta| b\alpha a f(\tilde{\mathbf{p}}_B) = f(\tilde{\mathbf{p}}_A) \tag{98}$$

This can be rearranged to give (93).

The case where α is negative and β is positive is similar to the last case. The case where both α and β are negative is non-physical since then $\alpha\mathbf{p}_A + \beta\mathbf{p}_B$ would have negative components. We see that (93) holds whenever the arguments of f in each term correspond to physical states. If these arguments do not all correspond to physical states then the equation does not correspond to any physical situation. For mathematical simplicity we will impose that (93) still holds in such cases.

Appendix 2

Proof of formula (58). We can write

$$\mathbf{v} = \begin{pmatrix} v_0 \\ v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} r_1 \\ r_2 - r_1 \\ r_3 \\ r_4 \end{pmatrix} \tag{99}$$

Hence,

$$\mathbf{r} = C\mathbf{v} \tag{100}$$

where

$$C = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{101}$$

Hence $\mathbf{r}^T D\mathbf{r}' = \mathbf{v}^T C^T DC\mathbf{v}'$. From (57) we obtain

$$F \equiv C^T DC = \begin{pmatrix} 2 & 1 & 1 & 1 \\ 1 & 1 & a & b \\ 1 & a' & 1 & c \\ 1 & b' & c' & 1 \end{pmatrix} \tag{102}$$

Now, $\mathbf{r}^I = \mathbf{r}_1 + \mathbf{r}_2 = (1, 1, 0, 0)^T$ (this is true for both the state vector (subscript S) and the measurement vector (subscript M)). The corresponding $\mathbf{v}$ type vector is, using (99), $\mathbf{v}^I = (1,0,0,0)^T$. Since $\mathbf{r}$ (which we take to be a M type vector) is normalized to μ then so is $\mathbf{v}$ and hence

$$\mu = \mathbf{v}^T F \mathbf{v}^I = 2v_0 + \sum_{i=1}^{3} v_i \tag{103}$$

Likewise, $\mathbf{v}'$ (taken to be a S type vector) is normalized to ν and so we have

$$\nu = (\mathbf{v}^I)^T F \mathbf{v}' = 2v'_0 + \sum_{i=1}^{3} v'_i \tag{104}$$

If $p_{\text{meas}} = \mathbf{v}^T F \mathbf{v}'$ is multiplied out and (103, 104) are used to eliminate v_0 then we obtain equation (58) as required.

Appendix 3

In this appendix we show that the matrix A must be symmetric (and hence D must also be symmetric). A transformation can be described by a matrix $Z \in \Gamma$ which acts on $\mathbf{p}$-type vector or equivalently by a matrix $X \in \Lambda$ which acts on a $\mathbf{r}$-type vector. Consider a small reversible transformation $X = 1 + \epsilon Y$ (such transformations must exist by Axiom 5). Since this transformation is reversible the inverse X^{-1} exists. We have (to first order in ϵ) $X^{-1} = 1 - \epsilon Y \in \Lambda$. Therefore, if $\mathbf{r} \to X\mathbf{r} = \mathbf{r} + d\mathbf{r}$ is an allowed transformation then $\mathbf{r} \to X\mathbf{r} = \mathbf{r} - d\mathbf{r}$ must also be an allowed transformation. If $\mathbf{r} \in Q_{\text{pure}}$ then $\mathbf{r}^T D \mathbf{r} = 1$. We must have $\mathbf{r}^T D(\mathbf{r} \pm d\mathbf{r}) \leq 1$ and $(\mathbf{r} \pm d\mathbf{r})^T D\mathbf{r} \leq 1$. Hence

$$\mathbf{r}^T D d\mathbf{r} = (d\mathbf{r})^T D\mathbf{r} = 0 \tag{105}$$

or,

$$\vec{v}^T A d\vec{v} = d\vec{v} A \vec{v} = 0 \tag{106}$$

for reversible transformations. We can take the transpose of the second of these equations and subtract it from the first equation. We obtain

$$\vec{v} B d\vec{v} = 0 \tag{107}$$

where $B = A - A^T$ is an antisymmetric matrix ($B^T = -B$) and can therefore be written

$$B = \begin{pmatrix} 0 & h_1 & -h_2 \\ -h_1 & 0 & h_3 \\ h_2 & -h_3 & 0 \end{pmatrix} \tag{108}$$

Hence, (107) gives

$$\vec{v} \times \vec{h} \cdot d\vec{v} = 0 \tag{109}$$

where $\vec{h} = (h_1, h_2, h_3)$. Consider a particular $\vec{v}$ which is not parallel to $\vec{h}$. Equation (109) implies that either (i) $d\vec{v}$ must lie in the plane containing $\vec{h}$ and $\vec{v}$ or (ii) $\vec{h} = \vec{0}$. Consider (i) first. The plane will intersect with the set R_{pure} giving either a set of disconnected points or a one dimensional curve (since R_{pure} are the extremal points of a convex set). The case of a series of disconnected points violates the continuity axiom. Hence, we must consider the case of a one dimensional curve. There will be a certain invertible matrix X which, along with its inverse, generates infinitesimal transformations along this curve. There cannot be any other infinitesimal reversible transformations since these would have the effect of moving points on this one dimensional curve off it violating (109). However, this one dimensional curve is contained in a two dimensional plane and cannot pass through all the fiducial vectors. This means that it will not be possible to find reversible transformations that can transform between any two pure states in violation of Axiom 5. Hence, possibility (i) is ruled out. This leaves possibility (ii) which implies that $A = A^T$. Hence, A and D must be symmetrical.